

NCER Working Paper Series
Sportometrics Program

Evidence of referees' national favouritism in rugby

Lionel Page
Katie Page

Working Paper #62
October 2010

Evidence of referees' national favouritism in rugby

Discussion paper

Lionel Page

University of Westminster & University of Cambridge

National Centre for Econometric Research, Queensland, Australia

Katie Page

Department of Psychology, Heythrop College, University of London

Introduction

The present article reports evidence of national favouritism from professional referees in two major sports: Rugby League and Rugby Union. National favouritism can appear when a referee is in charge of a match where one team (and only one) is from his country. For fear of the risk of such favouritism, such situations are avoided in most major sports. In this study we study two specific competitions who depart from this “national neutrality” rule: the European Super League in Rugby League (and its second tier competition) and the Super 14 in Rugby Union. In both cases we find strong evidence that referees favour teams from their own nationality, in a way which has a large influence on match results.

For these two major competitions, the Super League and the Super 14, we compare how a team performs in situations where the referee both shares their nationality and in situations where the referee comes from a different nationality. We also analyse referees' decisions within matches (such as penalty and try decisions) in a Rugby League competition, the Championship (second tier below the Super League). In both Rugby League and Rugby Union we find strong evidence of national favouritism.

1 Research on referees' decision making processes

Referees' decisions often have to be made in very complex situations which may surpass the limited cognitive abilities of the human brain (Salmela 1978). For example, in tennis the visual perceptual system is limited in that it does not allow a differentiation between a ball on the line or one just outside it (Jendrusch 2002). In sports which have adopted video refereeing this difficulty is exemplified by the fact that video decisions may take several minutes involving frame-by-frame footage from different angles. Whilst in the

same situations without video footage, referees are expected to make decisions in split seconds.

In addition to this intrinsic task difficulty, referees have to make these decisions under both high time pressure and high social pressure. The time pressure component comes from the fact that referees are expected to make binding decisions in a very short amount of time. Long decision times may signal uncertainty on the part of the referee and thus undermine the credibility of the referee for the players and the spectators. The social pressure component arises because decisions are usually made in front of partisan crowds who judge and scrutinise the referees' decisions. The crowd is usually unevenly distributed with a higher proportion of supporters for the home team.

It is therefore to be expected that referees make mistakes. Numerous sports have incorporated this fact into the design of their sports, going from the inclusion of video refereeing to the possibility for teams or players to challenge referees' decisions (tennis, cricket). On the contrary, other sports have been reluctant to adopt any measures which could be seen as undermining the authority of the referee in the game. The most notable case is football, which pays the price for this choice with recurrent blatant mistakes with significant consequences from the hand of Maradona to the recent incident involving Thierry Henry and Ireland.

Whilst it is to be expected, and perhaps tolerated, that referees may make some mistakes, it is more important to detect if those mistakes tend to follow some systematic pattern. It has been one of the foci of research studies on professional referees to detect the systematic biases in their decision making processes. The most clearly identified bias is the bias in favour of the home team. Previous research in football has found that referees are more likely to favour the home team in their decision making concerning fouls and cards (Dawson, Dobson, Goddard, and Wilson 2007), but also in the amount of extra time allocated to the home team is higher when the home team is behind (allowing it to catch up) than when it is ahead, (Dohmen 2008, Garicano, Palacios-Huerta, and Prendergast 2005, Sutter and Kocher 2004). Several studies have shown that the home team bias is largely induced by the home crowd (Boyko, Boyko, and Boyko 2007, Page and Page 2009). The simple existence in a football stadium of a running track reduces the proximity with the crowd and the home bias of the referee as well (Buraimo, Forrest, and Simmons 2007, Scoppa 2008). The home advantage has also found to be referee specific with some referees giving a large home advantage and others a smaller one. In addition, the effect of the crowd on the home advantage depends on the referee's identity (Page and Page 2009), with it increasing for some referees when the size of the crowd increases while it stays roughly constant for other referees (who may be more resilient to social pressure). A possible explanation is that referees use crowd noise as a clue about the existence of a foul. Interestingly, Nevill, Balmer, and Williams (2002) found that referees judging actions from video footage were much less likely to advantage the home team when the sound was turned off than when it was turned on, which suggests that the crowd

noise plays at least some part in this phenomenon.

Beyond the home advantage effect, the time pressure element may cause other type of biases. A long decision making process or a continual request for touch judge advice would undermine the referee's credibility. For this reason, referees make quick judgements in uncertain situations without time to reflect on all of the available information. It is therefore possible that they use prior knowledge of teams and/or players to help make their decisions (Plessner and Haar 2006). It has, for example, been shown that the previous reputation of players influences referees decisions in basketball (Lehman and Reifman 1987), baseball (Rainey, Larsen, and Stephenson 1989) and football (Jones 2002).

Finally, one may wonder if referees may be influenced in their decision making by implicit preferences they may hold over the different competitors. Preferences are not directly observable, however in some situations the referee's identity may suggest a natural default preference between two teams/competitors. This is particularly the case when the referee shares the national/local identity of one of the teams/competitors. The question of national/local favouritism from referees has attracted a significant amount of attention from researchers. National favouritism has been found in numerous sports and, in particular, in subjectively judged sports, with clear national favouritism found in figure skating (Seltzer and Glass 1991, Whissell, Lyons, Wilkinson, and Whissell 1993, Campbell and Galbraith 1996), gymnastics (Ansorge and Scheer 1988, Ste-Marie 1996), ski jumping (Zitzewitz 2006), rhythmic gymnastics (Popović 2000) and synchronised diving (Emerson, Seltzer, and Lin 2009). In all these sports, referees have been found to be systematically advantaging competitors from their own countries. The effect of this favouritism is significant in these competitions given the importance of referees' decisions on the final rankings.

In other sports where the results depends less directly on referees' decisions, some evidence has also been put forward about the existence of a national/local bias. A study on Australian football found that referees were more likely to favour teams from their own states in matches against a team from another state (Mohr and Larsen 1998). In cricket, (Crowe and Middelcorp 1996) found that Australian referees seem to make more leg before wicket (LBW) decisions in favour of the Australian teams in matches played in Australia over the period 1977-1994¹. More recently a study on MuayThai (Thai Kickboxing) (Myers, Balmer, Nevill, and Al-Nakeeb 2006) also found evidence of national favouritism from the judges.

In major sports like football and rugby, no such favouritism has yet been found. The study of such a topic is made difficult by the fact that these sports do not usually allow referees to officiate in matches with one team from their own nationality. In this paper we look at rugby competitions where referees' decisions are thought to have a rather limited role in the final results of the games (in that they are enforcing rules rather than judging a performance). We studied club competitions where referees officiate with one team being

¹They cannot eliminate other possible reasons beside favouritism to explain this pattern.

from their own nationality. We find a systematic national favouritism from referees, with a strong impact on the result of the matches. We discuss at the end of the paper the implications of these results for the major competitions in rugby and football. We also discuss several solutions to limit the effect of this favouritism on such competitions.

2 Data

2.1 Data on Match Results

Our dataset contains information on match results from two competitions in two different rugby codes. First, we have information from the European Super League, which is a rugby league competition consisting primarily of British teams with one French team, in the period from 2006 to 2009. Second, we have the results of the Super 14 for the 2009 season. This is a rugby union competition which consists of five South African teams, five New Zealand teams, and four Australian teams in an international club competition. In the observed period (2006 - 2009), these competitions have the following attributes:

- they contain teams of different nationalities;
- they have referees of different nationalities;
- and, most importantly for our analysis, they allow referees to officiate matches where the team of their own nationality is opposed to a team from another nationality.

This last feature is quite rare in the sporting world. Most international sports competitions specifically avoid selecting a referee of the same nationality as one of the teams (i.e. World Cup). A special feature of the aforementioned competitions is that they are club (not country) competitions where, arguably, nationality may seem a less salient element of the contest. They are also standard year-long competitions which are comparable with national competitions in which the local origin of the referee is usually not of import. These two features may provide a rationale for why, in these competitions, referees are permitted to officiate matches of teams from their own nationality.

Prior to 2009 in the Super 14 competition referees did not officiate matches where there was a conflict of nationality. Specifically, the referee was either of a different nationality to both teams or the referee was the same nationality of both teams (i.e. English referee with two English teams). The decision to allow referees to officiate games where only one team is from their nationality and the other is from a different nationality appears in 2009.

In the case of the Super League the competition was strictly composed of English teams in 2005, and when the French team was included it received the same “treatment” as other teams with the referees being picked from the pool of usual referees. Interestingly, over the period 2006-2009 the competition had referees of different nationalities with

mainly English referees, some Australian referees and one French referee. This variation in referees nationality was not present over the period 1996-1997 when a French team was also included in the competition.

2.2 Data on Referees's Decisions

In addition to studying the outcomes of these matches we also collected information on specific decisions from video footage taken from a lower league of rugby league (Championship), which, since 2009, includes a French team as well. These matches have the advantage not to have been broadcast on television (we discuss why this is interesting later on). Our data includes actions from 11 matches (55% of the season of the French team), totalising **220** actions. One match presents the interest of being refereed by a French referee while all the others are refereed by an English referee

2.3 Empirical Strategy

An analysis of referee decision making involves several difficulties. One of these is when studying a possible slight in referees decisions, a natural problem is to define what is a “correct decision”. While some match situations indicate obvious decisions, lots of situations involve a margin of interpretation (which is even more so when one takes into account that referees do not make decisions in slow motion replay). It is therefore understandable that referees may make “mistakes”, that is, make a decision contrary to what would be suggested by a slow motion replay. It is also understandable that fans who are prone to wishful thinking find themselves disagreeing on numerous occasion with the referee's decision, given the lack of clarity of the situation being judged.

To study referee's favouritism bias we will therefore use two kinds of strategies. First, we look for control groups where the neutrality of the referees in the game is more evident. Such control groups can be used to observe how the alleged non neutrality of the referee impacts on the result of the game. In particular, we test if the non neutrality of the game tends to affect the average results of the games² Second, we model how a referee should influence the game if he has a preference for one team. This allows us to compare the decision pattern of non neutral referees to the pattern they should adopt if they have a preference for one team which influences their decisions.

²Obviously it is possible that referees favour one team without this favouritism affecting significantly the final result, but arguably such a situation would inconsequential.

3 Model

Suppose an agent is officiating in a match where he has a preference for one team (Team A). The agent will aim to maximise the Probability of winning of Team A³.

In order to advantage Team A the referee can make biased decisions.

Definition 1. *A biased decision is a decision which is different from the decision the agent would have made in the same situation if he had no preference over the two teams competing.*

Several times during a match, *decision situations* (DS) occur. In a match, a DS is a situation where a referee may make three type of choices: a decision in favour of team A, a decision in favour of team B, or no decision. For simplicity, we will only consider here that in each DS, the agent must make a decision between Team A and Team B. Each DS is characterised by a level of ambiguity. Figure 1 shows a way to represent this idea. Any situation is characterised by a level of ambiguity $a \in [0, 1]$. When the evidence is weak the ambiguity is high, and at the limit, the decision could go either way. When the evidence is more clearly in favour of one team, the ambiguity in the situation is lower. When an unbiased referee observes a DS, the balance of evidence will determine which team the decision should go for. If the referee favours Team A, he has the option to switch some decisions he would have made for Team B to Team A. Referees are evaluated on the accuracy of their decisions by their sporting body and by the spectators, so any biased decisions have a cost. We will suppose that the cost of a biased decision is negatively correlated with ambiguity such that the more ambiguous a situation is, the less costly it is for the referee to make a biased decision, and the less ambiguous it is the more costly it is to make a biased decision. The ambiguity is observed at each period, before making the decision.

Let us call d a decision variable in a DS which takes the value 1 if the decision is biased in favour of Team A, 0 if the decision is not biased, and -1 if the decision is biased in favour of Team B. Let P be the probability of winning the match for Team A, and C the cost of the decision, the programme of the agent in each DS is:

$$\max_d P(d) - c(d)$$

For simplicity we will consider that a match is decomposed into T periods and that in each period one and only one DS occurs. During each period t a scoreline difference Δ_t is generated by the opposition between the two teams. For simplicity we will assume that this score difference is a continuous variable and that $\Delta_t \sim \mathcal{N}(\phi, \sigma)$ with ϕ is a parameter of difference in ability between the two teams.

³We do not need to assume here that this favouritism is conscious. Favouritism may well be partially or fully unconscious. We discuss this point later in the discussion section

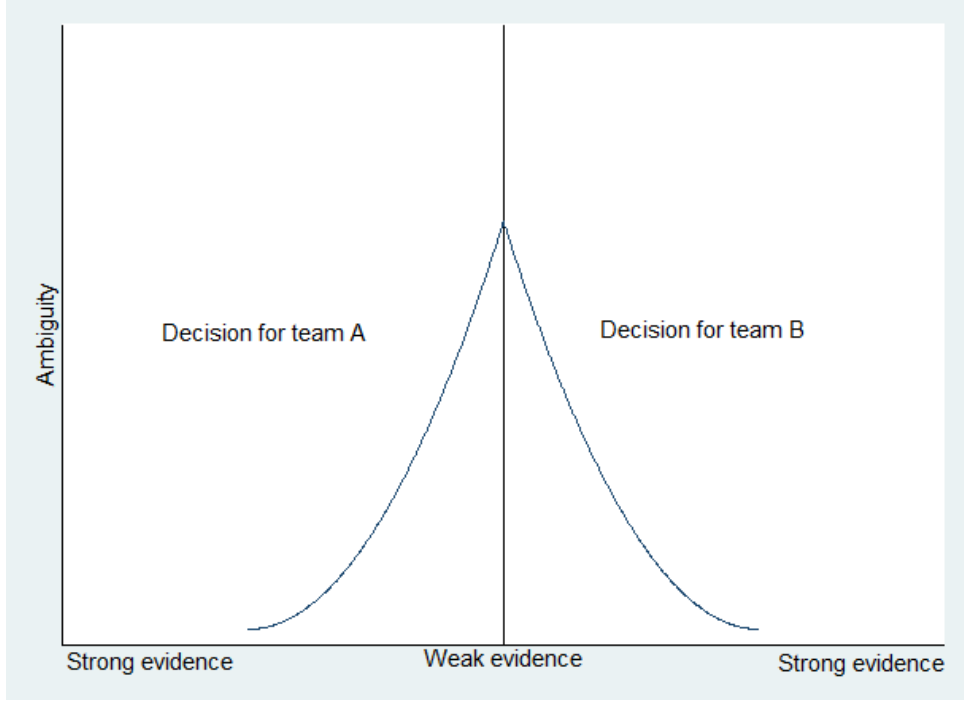


Figure 1: Decision situations and ambiguity

Let $\theta = T - t$ be the number of periods left to play. At any point in time the distribution of the total difference in points which will be seen later in the match is:

$$\Delta_\theta = \sum_{k=t+1}^T \Delta_k \rightsquigarrow \mathcal{N}(\theta\phi, \theta\sigma)$$

At each period t , the agent can decide to make a biased decision in order to influence Δ_{t+1} . A biased decision increases the expectation of the score difference for the current period for Team A. We will assume that a biased decision has an effect β_d on the scoreline difference at the next period, such that: $\Delta_{t+1}(d) \rightsquigarrow \mathcal{N}(\phi + \beta_d, \sigma)$ and as a consequence, $\Delta_\theta(d) \rightsquigarrow \mathcal{N}(\theta\phi + \beta_d, \theta\sigma)$.

Calling Δ_F the final scoreline at the end of the match, the unbiased probability⁴ of victory for Team A at any point in time in the match is: $P(\Delta_F > 0) = P(\Delta_t + \Delta_\theta > 0) = P(\Delta_\theta > -\Delta_t) = 1 - \Phi\left(\frac{-\Delta_t - \theta\phi}{\theta\sigma}\right)$.

The winning probability taking into account the possibility of a biased choice at the present period is therefore⁵

⁴The probability of the team winning if no manipulation has occurred or no bias is present.

⁵We purposely chose to assume here that the agent does not anticipate his possible biased decisions later on and take the unbiased probability as the decision variable. A simple way to relax this hypothesis

$$P(\Delta_t, \theta, d) = 1 - \Phi\left(\frac{-\Delta_t - \theta\phi - \beta_d}{\theta\sigma}\right)$$

Biased decisions have a cost $c(d)$. We will first assume that there is a direct cost $1 - a$ with a being the level of ambiguity of the DS. We will also assume that at the end of the match, the referee has a preference for a small difference in the number of decisions given between the two teams, $|\sum_t d_t|$. The reason for this is that a large difference could signal favouritism. The agent will therefore prefer not to create too large a difference between the number of decisions given for one team and those given for the other team, over the whole match. Let ψ be the unit cost of a difference in the number of decisions at the end of the match. At any point in time the total cost for all the decisions taken in the match is: $C = \sum_t a_t 1_{d_t \neq 0} - \psi(|\sum_t d_t|)$.

The marginal cost of a biased decision is therefore:

$$c_t = a_t 1_{d_t \neq 0} + \psi \left(1_{d_t \sum_{i=1}^{t-1} d_i > 0} - 1_{d_t \sum_{i=1}^{t-1} d_i < 0} \right)$$

The programme of the decision maker at a given period t is:

$$\max_{d_t} P(\Delta_t, \theta, d_t) - c_t(d_t)$$

A biased decision is taken in favour of Team A if:

$$\Phi\left(\frac{-\Delta_t - \theta\phi}{\theta\sigma}\right) - \Phi\left(\frac{-\Delta_t - \theta\phi - \beta_{d_t}}{\theta\sigma}\right) > a_t + \psi \left(1_{d_t \sum_{i=1}^{t-1} d_i > 0} - 1_{d_t \sum_{i=1}^{t-1} d_i < 0} \right) \quad (1)$$

Proposition 1 (Biased decisions). *Assuming that the probability of team A and team B to commit an offence or an error are similar, if a referee favours team A, the expected number of decisions over a match in favour of team A will exceed the expected number of decisions made for team B.*

Proposition 1 is intuitive, though it can only a result on average as a biased decision can only occur in so far as the level of ambiguity in a situation is high enough. It is therefore well possible in some matches that even a referee with favouritism give a penalty count against the favoured team because there were not enough ambiguous situations for biased decisions to compensate obvious decisions.

Proposition 2 (Importance of Decisions). *If the cost of different decisions can be considered constant, favouritism will induce more biased decisions for decision with a greater impact on the final result.*

would be to assume that the agent expects an average level of bias during the rest of the game, this could be entered in the model as affecting ϕ , with $\phi' > \phi$ being more in favour of Team A. An agent foreseeing perfectly the future would anticipate that his future choices to bias decisions will be affected by the future situations likely to be observed after his present choices.

If the cost of more important decisions for the referee is lower than the additional gains for Team A, critical decisions will tend to be more biased than less critical decisions.

Proposition 3 (Scoreline Difference). *When teams are close in ability, favouritism will induce more biased decisions when the scoreline is close and when the favoured team is just trailing.*

Proposition 3 (proofs in the Appendix) is very intuitive, it shows that the action of a referee is most likely to have an impact on the end result of the game when the biased decision is taken when the scoreline is close. In addition, it shows that in a DS where a biased decision is most likely to be taken are situations where Team A is slightly trailing (by half the average effect of a referee decision). Intuitively, it is when a team is slightly trailing that a biased decision will be the most helpful to help the team in a difficult situation. If the team is slightly ahead it does not need as many biased decisions to win the game eventually.

Proposition 4 (Scrutiny and Ambiguity of Decisions Situations). *Favouritism will induce more biased decisions when the ambiguity in the situation is larger.*

This proposition is straightforward. The ambiguity can be seen not just as something perceived by the referee, but how the referee perceives that other actors perceive the ambiguity. For example, if a large screen is present in the stadium and replays in slow motion the actions, for a given action, there will be less ambiguity for the spectators about the best action to be taken, hence the cost of a biased decision will be higher.

Corollary 1. *There will be less favouritism in DS facing a stronger scrutiny from the audience like the matches broadcast on television: replays in the stadium on a large screen, video refereeing decisions, replays and analyses of decision by television commentators.*

Proposition 5 (Scrutiny on the Balance of Decisions). *Favouritism will induce less biased decisions when the balance of decision is the object of a stronger scrutiny with ψ being higher.*

This proposition is straightforward. When the cost of creating an unbalance in the decision is higher, the favouritism will be lower.

Proposition 6 (Time Remaining in the Match).

- i When the scoreline is close favouritism will induce more biased decision toward the end of the match.*
- ii When the scoreline difference is large, favouritism will induce less biased decision toward the end of the match.*

Proposition 6 expresses the fact that the time left in the game affects the uncertainty of the final results. When the scoreline is close, if there is a significant amount of time left until the end of the game, the effect of a biased decision can be compensated by the normal evolution of the scoreline afterward. On the contrary, when there is not much time left in the match the effect of a biased decision may be even greater on the winning probability as there will not be much time left for Team B to score again.

When the scoreline is not close, if there is a significant amount of time left until the end of the game, the evolution of the scoreline later on may allow Team A to eventually win, and the biased decision would have been one of the elements contributing to this victory. On the contrary, if there is not much time left, a biased decision will not have any effect on the final outcome as Team A will not have enough time to score again and win the game.

Proposition 6 is very useful to identify favouritism because of a specific timing pattern which is asymmetric for Team A and Team B.

Proposition 7 (Compensation). *Favouritism will induce biased decisions for Team B in order to limit the differences in the number of decisions. Referees will compensate in favour of Team B when the effect of the final result is likely to be small.*

If during the match biased decisions are taken at the critical moments identified by Proposition 3, and 6, the balance of decisions will look uneven at the end of the game. Biased decisions in favour of Team B can then be optimal in DS where they would have a very low impact on the probability of Team B to win the match if they allow the referee to even the balance of decisions.

Proposition 8 (Decisions Marginal Value). *Favouritism will induce an asymmetry between the marginal value of referees' decisions:*

- i Biased decisions in favour of Team A on average produce more points for this team than a fair decision for Team A.*
- ii Biased decisions in favour of Team A on average produce more points for this team than decisions for Team B do for that team.*
- iii Differences in the penalty count creates larger differences in scores in matches with biased decisions.*

This conjecture is a simple consequence of the fact that decisions in favour of Team A tend to be taken when their impact on the game tends to be the greatest while decisions in favour of Team B tend to be taken when the impact on the game tend to be small. This conjecture cannot be proven without any assumption on the distribution of the actual scoreline during the match (a proof is given in the Appendix for a special case).

4 Results

4.1 Overall match results as a function of the referee’s nationality

Looking at the overall match results, we can compare how a team performs depending on the neutrality of the referee. In the Super 14 Rugby, we can take the home team as a reference and compare how they perform depending on the nationality of the referee. In the European Super League, there is only one non British team, so in this case we use this team as a reference and compare it’s results as a function of the nationality of the referee.

Figure 2 shows the average difference in score in favour of the home team and the winning probability of the home team for both of these situations. In both leagues, teams perform better when the referee is neutral or from their own nationality. The result is only significant for the proportion of wins in Super League with the Super League winning 67% of their match with an Australian or French referee, but only 41% of their matches with an English referee.

One may be worried that referees may be allocated to different kinds of game. For instance, in Super League, if Australian referees were to have been allocated to matches where the opposition to the French team is of lower quality, the comparison between neutral and non neutral referees would give the impression that the French team is more successful with a neutral referee. In Super 14 if the a referee from the nationality of the opposition is allocated when home team meet stronger opposition, the nationality could also hide a difference in the quality of the teams.

To control for this possibility, we can focus on matches between comparable opposition. In the Super 14, we restricted the sample to matches where the home team meets a team whose final difference in ranks in the championship is not higher than 3. For the Super League, we used the fact that we observe teams over several years to keep “fixtures” (home team-away team) where both neutral and non neutral referees have officiated.

It is noticeable that in both the Super 14 and the Super League, the “referee effect” is stronger in the subsample of comparable matches. The differences are now significant for the differences in terms of wins and points in both the Super 14 and the Super League.

In both situations the increase in the magnitude of the differences is actually due to the fact that teams are more likely to play in a match with a referee having the nationality of the opposite team when the opposition is weaker. That is why controlling for the differences in teams quality across matches reveals a larger effect of having a non neutral referee.

To go beyond these descriptive statistics, we implement within estimators which look at the the effect of the referee’s non neutrality by comparing the results of the teams with a neutral and non-neutral referee only within each set of comparable matches.

Table 1 shows the results of the statistical analyses. The variable “Referee proximity”

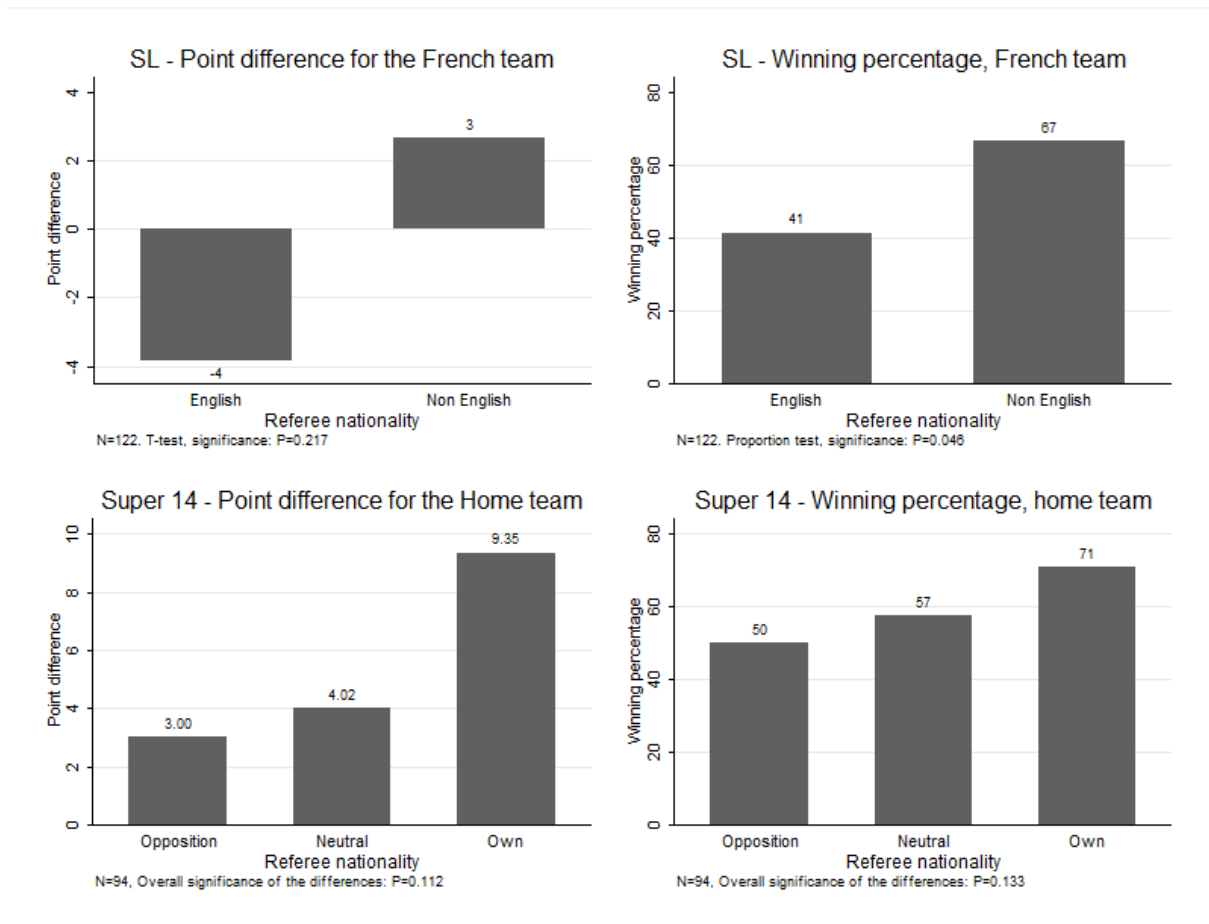


Figure 2: Team results as a function of the referee's nationality, Super 14 and Super League

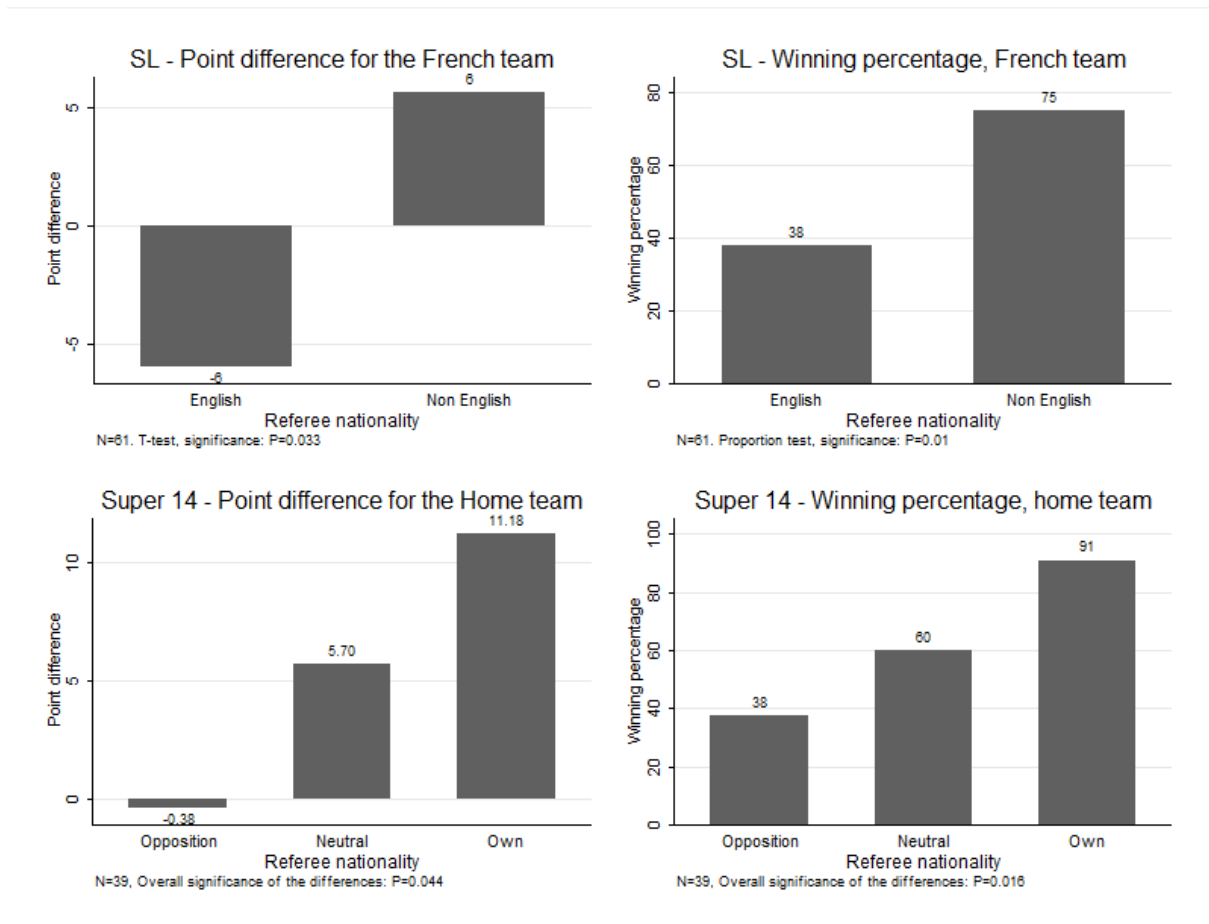


Figure 3: Team results as a function of the referee's nationality, Super 14 and Super League - Subsample of comparable matches

takes the value -1, 0 and 1 for the Super 14 sample to indicate respectively a referee nationality identical to the opposition team, neutral and identical to the home team observed differences in Super League are significant. In Super League, the French referee and the Australian referees are merged in one category and the variable takes the value 0 if the referee is English and 1 if he is not English.

Table 1: Referees national favouritism, sample of similar matches

| | Super 14 | | Super League | |
|-------------------|------------------------------|------------------|------------------------------|------------------|
| | FE Linear Δ points | FE Logit Win | FE Linear Δ points | FE Logit Win |
| | (1) | (2) | (3) | (4) |
| Referee proximity | 5.29 [†] (1.99) | 1.12* (1.97) | 9.07 [†] (1.72) | 1.27* (2.28) |
| Constant | 5.59* (2.69) | | -5.26** (-1.88) | |
| N | 27 | 27 | 57 | 57 |

Robust t-stats and z-stats in parenthesis, significant at [†] 10%, * 5%, ** 1%

Table 1 shows that even with few observations when keeping only comparable matches, and even when using a within estimator, the effect of the referee nationality is large and statistically significant in both competitions. In Super 14, a referee with the nationality of the team increases the scoreline of the team by 5 points relative to neutrality. In Super League, the French team received, on average, 9 points more in a match when the referee is not English.

In the case of favouritism, Corollary 1 suggests that referees should influence the game less when it is broadcast on television. We collected the list of matches which were broadcast live on English television for the Super League over the seasons 2006-2009. Figure 4 shows the difference of the French team results whether the game is broadcast live on television or not. We only used comparable fixtures, that is fixtures which have sometimes been on television and sometimes not. The difference is surprisingly very strong. When the referee is English, the French team is much more successful when the match is on television (59%) than when it is not on television (30%).

Here again we can control for differences between the fixture quality using a fixed effect estimator. Table 2 shows these results.

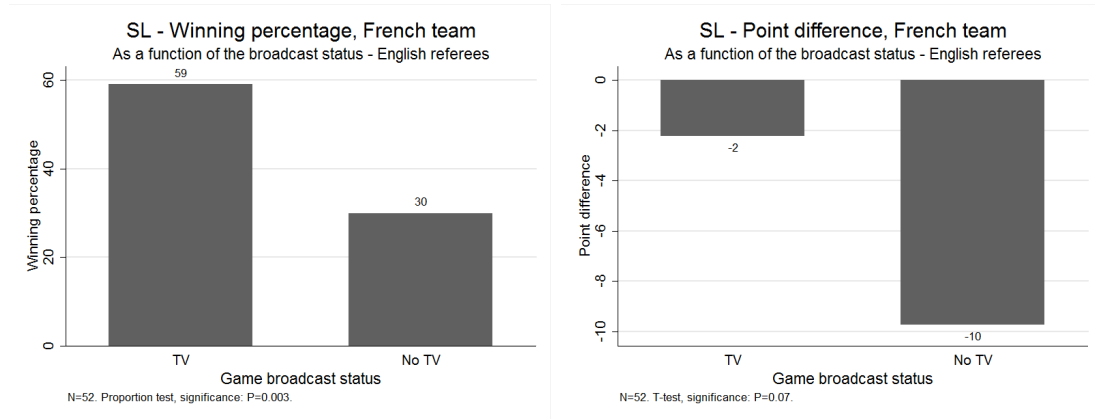


Figure 4: Results of the French team in Super League depending on broadcasting status of the match, when the referee is English.

Table 2: French team results with English referees, effect of live broadcasting

| | Super League | |
|------------|------------------------------|------------------|
| | FE Linear Δ points | FE Logit Win |
| | (1) | (2) |
| Live on TV | 9.78 [†] (1.90) | 1.84* (2.20) |
| Constant | -13.04* (-4.45) | |
| N | 52 | 52 |

Robust t-stats and z-stats in parenthesis,
significant at [†] 10%, * 5%, ** 1%

4.2 Referees' decisions within matches as a function of their nationality

The first section explored the impact of referee decisions on match outcomes. In this section we explore in more depth within-match decisions in order to shed more light on the specific nature of the favouritism bias. From the match records of the Super League, we compare the number of penalties given to each team as a function of the nationality of the referee. Figure 5 shows that English referees award significantly more penalties to the English team ($P < 0.01$) while this is not the case for non-English referees ($P = 0.81$).

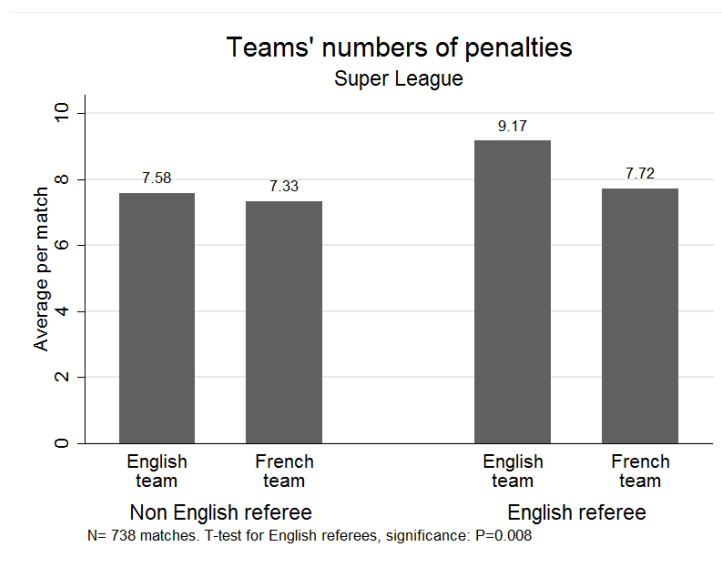


Figure 5: Penalties awarded as a function of the referee's nationality

This information is however limited to penalties and does not take into account scrum decisions. We analysed the video footage from 11 matches involving a French team in the English Championship (second tier competition of the Super League).⁶ We find that English referees are more likely to make decisions (penalties and scrums) in favour of the English team, as the first two bars of Figure 6 show. In the only match with a French referee, played in England, the proportion of decisions in favour of the English team (53%) was much smaller than for matches played in England with English referees (59%).⁷

⁶These 11 matches represents 55% of their regular season. Interestingly, most video footage was produced by the clubs, and only one of the matches was broadcast live on English television. As we found that favouritism is more likely to be present in matches not broadcast live on television, these are video footage of the situations where a favouritism is most likely to be seen.

⁷This comparison must however be taken with caution because it relies on only one match featuring a French referee.

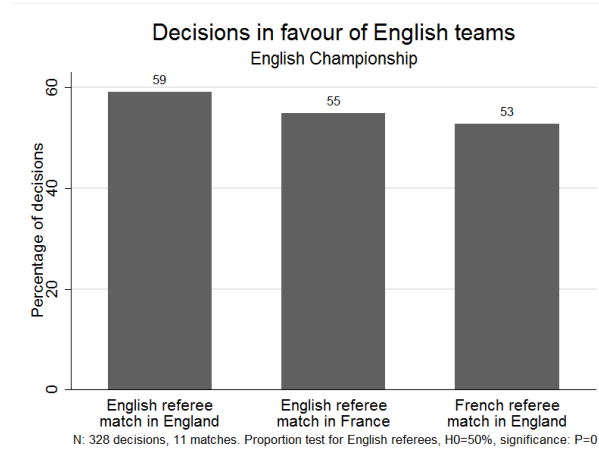


Figure 6: Referees' decisions as a function of their nationality and of the location of the match

4.2.1 Favouritism is particularly strong for critical decisions (Proposition 2)

Beyond the simple proportion of decisions in favour of each team, particular decisions are more likely to go the way of the team from the referee's nationality when they are critical. Figure 7 shows that for non-English referees there is little difference with respect to the number of cards given to the French and English teams⁸. On the contrary, in matches where an English referee officiates, the French team receives more than twice the number of cards of the English team. Relative to non-English referees, English referees penalise the English team less and the French team more. The number of yellow cards given by English referees is significantly higher for the French team than for the English teams ($P=0.03$).

Figure 8 shows that a similar pattern exists in the English Championship with the French team receiving a higher amount of yellow cards than the English teams. Whilst we do not have a control group with non English referees in the Championship, it is possible to observe that compared with matches without any French team most of the difference comes from the fact that English teams are less likely to be sanctioned by yellow cards when they play against a French team ($P=0.057$). A similar number of cards is given to the French team and to an English team in all-English matches implying that poorer discipline on the part of the French team is unlikely to be the cause of the observed

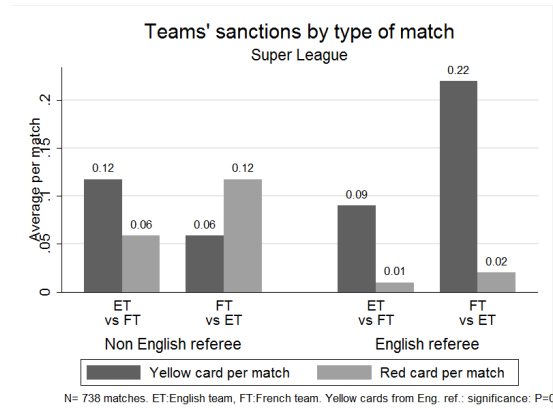


Figure 7: Sanctions in Super League

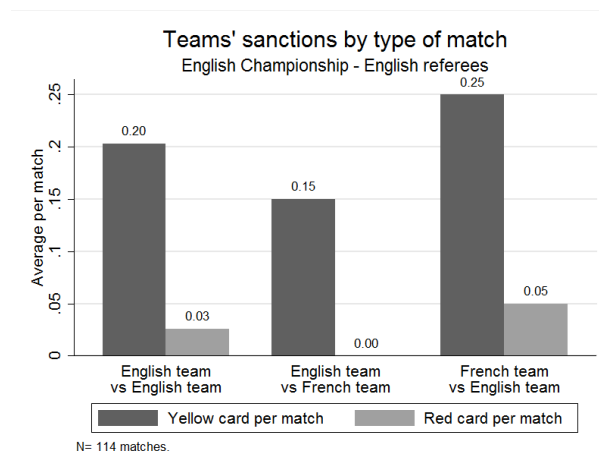


Figure 8: Sanctions in Championship

differences.⁹

Interestingly, for the Championship we were also able to look at the critical decisions to validate tries when the ball has been grounded. We found that the French team received a lower proportion of positive decisions regarding try validation (79% vs 93%). The difference in proportion of tries validated is statistically significant.

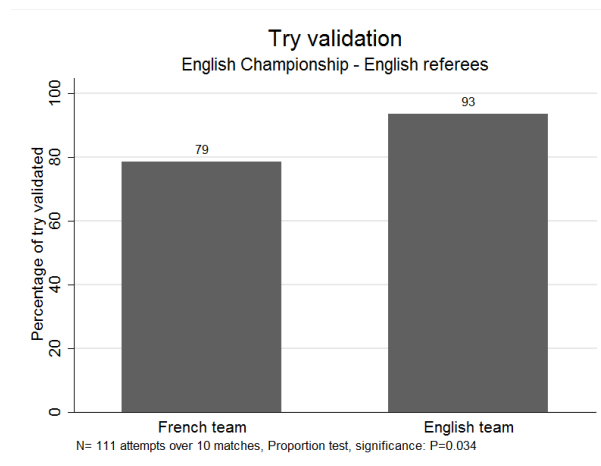


Figure 9: Proportion of tries validated

4.2.2 Favouritism is particularly strong when the scoreline is close (Proposition 3)

In addition to the types of sanctions, the timing of the decisions also suggests national favouritism. English referees are much more likely to make decisions in favour of the English team when the score is close. Changes in scores are the most likely to impact on the final result when the game is close¹⁰. Therefore, the difference observed in Figure 10 in the proportion of decisions taken by the English and French referees masks a more fundamental difference in their timing. In situations with close scorelines (0 to 4 points difference) the English referees only gave 58% of the decision in favour of the English team versus only 43% for the French referee. Decisions by English referees favoured more the

⁸French teams get more red cards and less yellow cards. However, given the small number of occurrences no inference can be made from these numbers.

⁹Overall, the number of sanctions (penalties and cards) is actually lower in matches with a French team than in matches between two English teams. This observation also points to the fact that the difference between French and English teams is not due to poorer discipline from the French team. It is just the allocation of these sanctions which creates a difference between the French and English teams.

¹⁰A decision one way or the other when the game is close may help a team to get ahead and possibly build on this advantage. When there is a significant difference in scoreline a decision is not as likely to have an impact on the final result.

English teams when it mattered the most, and symmetrically, decisions from the French referee favoured more the French team when it mattered the most.

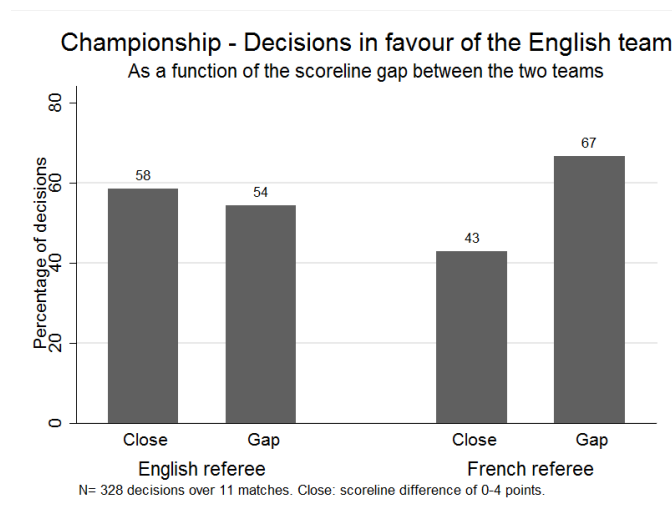


Figure 10: Timing of referees decisions

While there are few critical decisions, their timing is also paramount given their importance. We find that when the scoreline is close the difference between the proportion of tries validated for the French team and for the English team increases such that the French team has only a 59% chance to get a try validated versus 82% for the English team¹¹. Figure 11 shows that when the scoreline is close both teams have a lower try validation which may be due to a higher difficulty to score cleanly in a close contest. However, the gap for the French team is much larger (22 percentage points) when the scoreline is close than when it is not (14 percentage points).

¹¹We did not include here the television match as the try validation is often referred to a video referee.

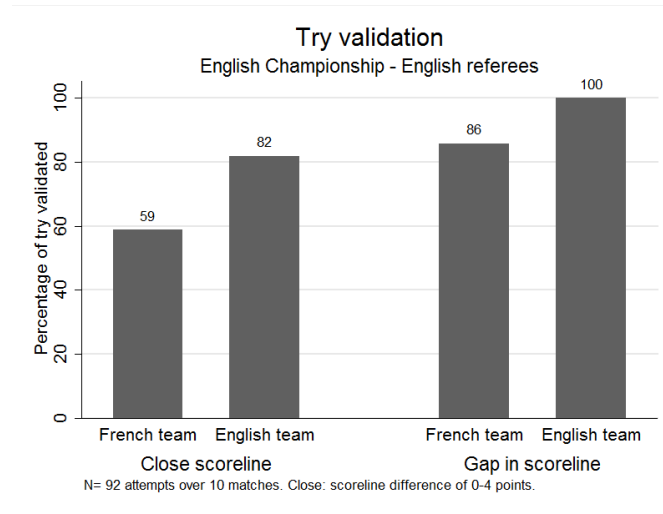


Figure 11: Timing of critical decisions

4.2.3 Referees' decisions are more in favour of their national team when the level of scrutiny is lower (Propositions 4 and 5)

The level of scrutiny for English referees should be minimal in matches in England which are not broadcast on television. Figure 12 shows that it is in situations of greater scrutiny that the timing of the decisions are the most asymmetric with 80% of the decisions going the way of the English team when the scoreline is close. While this number is only based on a very limited number of decisions (10 decisions only), it goes some way to explaining the 27% proportion of wins for the French team in SL when the game is not broadcast on television.

Finally, looking at the type of referees' decisions, it is possible to consider that the level of scrutiny is not the same for every decisions. Penalties are recorded and often published with the match results. On the contrary decisions to allocate the ball in scrums following knock-ons or forward passes are not recorded. For this reason the referee has potentially more discretion in these choices which face a lower level of scrutiny. Figure 13 shows that indeed, English referees in the Championship make more decisions in favour of the English team for scrum decisions than for penalty decisions (this difference is however not significant).

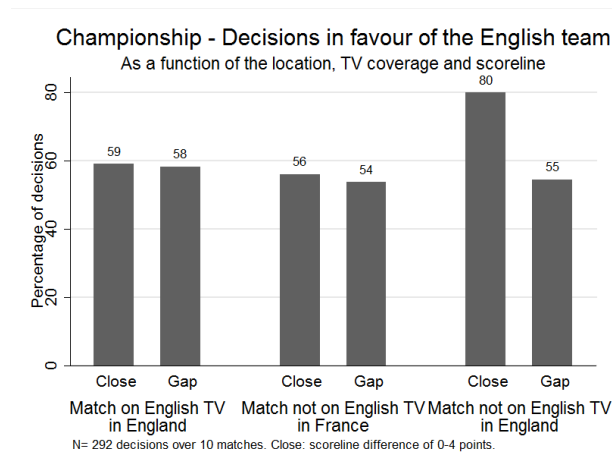


Figure 12: Timing of the decisions as a function of the location and broadcast status - English referees only

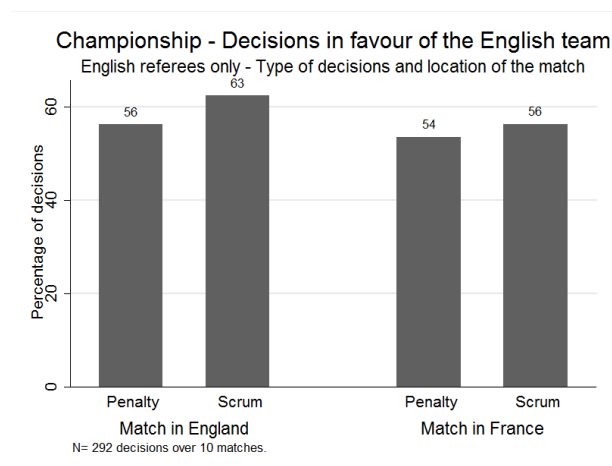


Figure 13: Proportion of penalty and scrum decisions taken for each team

4.2.4 When the score is close favouritism induces more biased decisions when there not much time left in the game. When the score is not close, favouritism induces less biased decisions when there is not much time left in the game (Propositions 6).

Figure 14 shows that when the scoreline is relatively close (12 points or less), the proportion of decisions in favour of the French team is lower at the end of match. On the contrary, when the scoreline difference is relatively large, more decisions are made for the French team at the end of the match. These differences are large in absolute terms, however they are not significant due to a small sample size and should only be considered as suggestive.

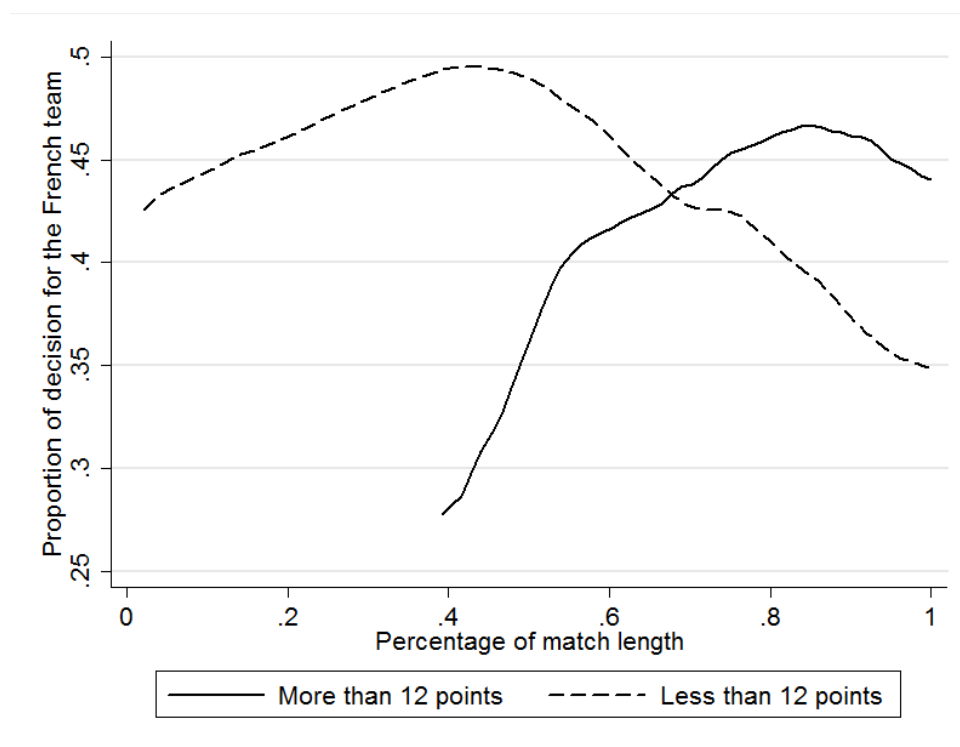


Figure 14: Decisions for the French team in the Championship as a function of the time and of the scoreline - English referees

4.2.5 In case of favouritism the average number of points generated by a decision for team A is larger than fair decisions (Proposition 8)

A simple way to test Proposition 8 is to see if on average the penalty count difference observed in matches with a non neutral referee creates a larger difference in points than in matches with a neutral referee. Table 3 presents the average marginal effect on the

| | Super League | Championship |
|------------|--------------|--------------|
| ET vs ET | 0.42 | 1.13 |
| FT vs ET | 1.96 | 1.71 |
| Difference | 1.53* | 0.59 |
| N | 394 | 97 |

Significant at: *, 5%.

ET: English team, FT: French team.

Table 3: Marginal value in points of penalty decisions accross types of matches - English referees

scoreline of penalty decisions. These marginal effects are computed by linear regression of the final scoreline difference on the difference in the penalty count. It shows that indeed, when English referees officiate in a match with a French team, the average difference in points created by a given difference in penalty count is larger than when the match opposes two English teams. In Super League, a difference of one penalty in favour of the the English team produces on average 1.53 points more when it is a match with a French team and an English referee. This difference is statistically significant ($P=0.024$). In the Championship penalties also produce more points in the match with the French team (this difference is however not significant).

5 Discussion

5.1 Summary of Findings

In the Super 14 competition (rugby union), the European Super League and the English Championship (rugby league) our study finds that teams are much more successful when the referee is from their own nationality. In the Super 14, the home team wins 71% of its matches when the referee is from its own nationality and only 50% when the referee is from the nationality of the away team. In the Super League, the French team wins 67% of its matches when the referee is not English (Australian or French) and only 41% when the referee is English. However, these differences fade when referees' decisions face a higher level of scrutiny. In the Super League matches where the referee is English, the French team is much more likely to win if the match is broadcast live on English television (60%) than if it is not (37%).

Our analysis of referees' decisions within matches indicates that referees make more decision in favour of the team of their own nationality and that, moreover, favouritism is more discernible on most critical decisions like yellow and red cards and tries validation. We also find evidence of referees making decisions in favour of the teams of their own at

the most critical times during the match (i.e. when the scoreline is close).

5.2 Conscious or unconscious biases?

A natural question arising from these results is how conscious/intentional these biases are. It is beyond the scope of this paper to answer this question, however we should stress here that this a difficult question to answer.

There are at least two arguments which can be put forward to defend that a least a significant part of these biases may be unintentional. First, referees' decisions are made under intense time pressure because referees have to make decisions in split seconds. Research on the Implicit Association Test (Greenwald, McGhee, Schwartz, et al. 1998, McConnell and Leibold 2001) has found that for such quick decisions, implicit preferences matter with individuals taking more time to make the right decision when it goes against their implicit preference than when it reinforces it¹². This could affect referees' decision processes if they have a threshold in terms of decision time to make a decision. Given the social pressure, referees may feel uncomfortable to make a decision one way or the other after a given time threshold following an action/incident (for instance after 3 seconds). Then, when a decision has to be made by a referee against a team for which he has implicit preferences, the time required to make the decision may be too long and referees may be more likely to be beyond the decision threshold in situations where the decision would go against the team they support. Figure 15 shows such possibility with the distribution of time the referee needs to make a decision in both cases. If the referee opts not to make any decision over a given threshold and if decisions against a given team take more time to be made, they will less likely to be made.

Research on cheating has also shown that the psychological border between the right and wrong decision is fuzzy (Ariely 2008). Most decisions require an interpretation of the rules to determine the best action to pursue. In most situations there is some ambiguity about the perfect interpretation of the rule. This ambiguity typically allows the individual to adopt self serving behaviours which he/she can rationalise as being in accordance with the rules. It has for instance been found that cheating behaviours are much more likely when there is potential ambiguity about the interpretation of the rules. Ambiguity is a major element in situations referees have to judge. In their 2002 study, Nevill, Balmer, and Williams (2002) proposed 47 football situations to a set of referees on video footage. Non of them received a unanimous verdict from referees. This suggests that in team sports like football, and certainly like rugby, there is a large margin of interpretation for the referee. In numerous situations, the best decision to enforce the rules is not straightforward¹³. In such ambiguous situations, the decisions may be systematically

¹²Implicit Association Test (IAT) is a test designed to measure implicit attitudes, that is attitudes which are outside of the individual's awareness.

¹³In rugby league for instance, it is often hard to visually determine if a ball has gone forward (knock on, forward pass) or not (ball in play), it is hard to see if a ball has been stolen in a tackle (ball steal)

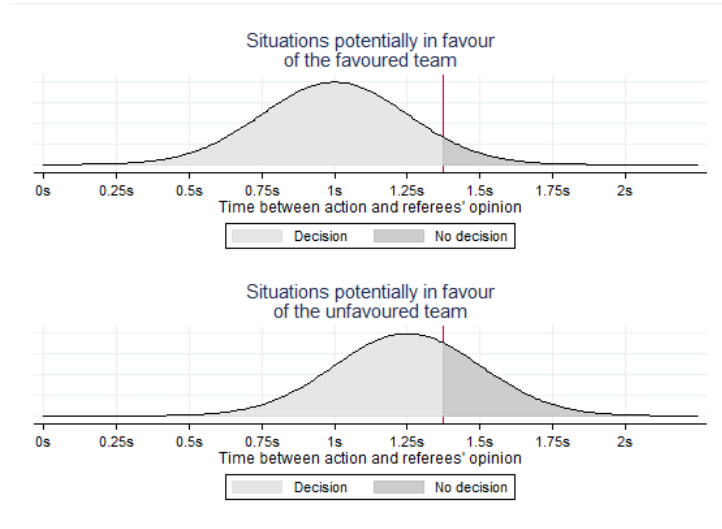


Figure 15: Timing of decisions and favouritism

biased in favour of the individuals' preferences (here the referee), while in the mean time the ambiguity offers enough room for them to rationalise their choices as being the right ones. In such situations the individual may be able to stay convinced that he has been following the right decision process while adopting a systematically lopsided decision process.

Part of this national favouritism could naturally be coming from an intentional desire to affect the result of the game¹⁴. Such a claim cannot be made or rejected from the data we observe. Whilst it is of interest for anybody willing to assess the level of responsibility (loosely defined) of the referees in this situation, it is however not our primary interest here. The level of consciousness of referees' biased decision making process is certainly an interesting psychological question. However, in a more economic approach (in both senses of the term) we simply show here that referees' national preferences do matter and that they impact significantly on the results of the competition.

5.3 Implications

These results are important in the whole set of research on referees' decision making. More practically they are important for sports competitions involving teams from different countries in rugby and in other sports. They suggest that the referee's identity cannot be ignored and that referees' preferences may well influence the game.

or if the player lost it (knock on).

¹⁴Such intentional biases have for instance been found in situations of corruption (Distaso, Leonida, Patti, and Navarra 2008). Such situations are, however, most likely to be rare.

For the competition under investigation in this paper the existence of a favouritism bias is problematic as it affects the fair outcome of the competition which is costly to both the teams and supporters. Beyond this, the results could also affect the market value of television rights for television channels from different countries because television channels will typically provide a higher value to matches where the national team is more likely to win. In the specific case of the Super League of Rugby the peculiar situation of a single French team mainly refereed by English officials suggests that favouritism bias may be a major obstacle to the goal of expanding the game in France, which is officially one of the reasons for the inclusion of the French team in the first place¹⁵.

Beyond these specific competitions our results suggest that national favouritism should not be ignored in major professional sports. More subtle favouritism like the possible preferences of referees over local teams in a national championship, or the possible preference over teams from one's own continent¹⁶ in international matches, or even the strategic preferences over a result in a game where the referee is neutral but where the result may have a consequence on the referee's national team prospects¹⁷.

These results may also be considered as having more interest than just the identification of a national favouritism bias. A major element of our results is that the differences across referees are very large, a result which indeed surprised us. This suggests that referees have a in rugby a very large impact on the results. Whilst we can measure here differences between referees on some observable characteristics, we do not observe differences which could arise from their preferences/beliefs of the players, their susceptibility to be influenced by the crowd, their beliefs in the specific way to interpret certain rules, or even their frame of mind on a given day including their expectation from the present match¹⁸. This shows that the quest for a fair and accurate refereeing is a goal which is not easy to achieve. Our results suggest that a sport like rugby under both its codes should try to continue to look into new solutions to improve the refereeing process and overcome its human limitations¹⁹.

¹⁵In the specific case of the Championship, a French team was included in this second tier competition as a stepping stone to be included in the Super League. The final inclusion is however conditional on a high level of performance in this second tier. Our results show that the systematic bias suffered by this team may well be a major factor preventing this team to perform and therefore paradoxically making a failure of this stepping stone strategy.

¹⁶When for instance a European referee is in charge an international game involving a European and a non European team.

¹⁷In the 2009 Six Nations Tournament in Rugby Union, a Welsh referee notably allocated a penalty count of 13 to 2 in favour of Ireland in a match Ireland versus France while France was at that stage the main contender for the bookmakers alongside Wales. Whilst there is no evidence that the referee was biased in his decision making process, such situations present clearly an interest as they may arise in numerous competitions even if the matches with non neutral referees are avoided.

¹⁸This potential over importance of idiosyncratic referees performance over the match was voiced loudly by the Saracens manager in 2010: "There is a problem and that is the game is determined by referees, not by teams. Everybody wants to know why the game is dying maybe that is why." *Brendan Venter*.

¹⁹The choice by the National Rugby League to use two referees would deserve to be studied as it

Finally, and possibly more importantly, this national favouritism in sport from professional referees is a result of a pure case of the effect of national preferences. Professional referees have an important material interest to make the best decision possible (as their career depends on it). On the contrary no material interest is associated with the allocation of an advantage to the team of their country²⁰. This national favouritism from referees is therefore a unique situation where individuals can be observed to act purely for national preferences. While national preferences have mostly been ignored in economic models of individual behaviour, our results show without a doubt that national identity does matter.

5.4 Solutions

First, the easiest solution to consider is to limit the number of potential contentious situations. For instance, in cricket, as a result of concerns over leg by wicket decisions by home referees, cricket authorities have suggested that each series is umpired by both a home and a neutral umpire (Crowe and Middelorp 1996). The introduction of the non neutral referees in Super 14 in 2009 may be seen on the contrary as going in the wrong direction. Similarly, the present absence of any Australian referees in Super League mechanically reduces the number of matches with neutral referees..

Second, another easy solution is to make referees aware of this situation as a part of their training. It is not clear how this may affect their decision making, but feedback to referees regarding their biases has been found to reduce their bias in other disciplines (Plessner and Haar 2006).

Third, a natural way to improve referees' decision making is to incorporate appropriate incentives to make adequate decisions. The highest the private incentives are the lower the influence of the national preferences are likely to be. As Rickman and Witt (2008) have shown, the home bias from referees significantly reduced with professionalism as referees performance got more closely monitored and used in their professional promotion. A simple solution may therefore be to increase the monitoring of non neutral refereeing to judge referees' performance. An increased monitoring of referees decision making in non neutral matches would also be likely to decrease potential biases. This could be done by having video refereeing at non neutral matches. Another possibility would be to transfer part of the monitoring process to the clubs themselves who could challenge referees decisions, if not during the game, like in tennis, possibly after the game. Clubs could be given several opportunities to challenge referees decisions after one match and to build a case based on video footage. These additional specific reviews would then be added to the evaluation of referees over the years. Along these lines another possibility

could intuitively lead to a lessening of individual biases (while on the other hand new biases from mutual influence could arguably arise).

²⁰It is reasonable to assume so.

would be to introduce the challenge option in the game itself with teams being allowed a limited number of challenges for some specific action. Decision would then be referred to video officials.

5.5 Using a fairness index as a screening tool

In order to help this additional screening, the model developed in the present paper actually provide a way to measure an index of fairness of matches and referees which may help sport organisation and clubs to identify which matches or referee needs additional screening.

A simple use of previous analysis is to compare the results from some categories of referees to some control group. For instance, from the same analysis used to compute Table 1 results it is possible to measure individual effect for some referees. While this is not possible for the Super 14 due a sample size too small. It is possible for the Super League for a limited number of referees with enough observations. Table ... shows the difference in score for the French team between some English referees and a control group composed uniquely of Australian referees. These numbers could be used as index of fairness for each referees.

Such an approach is however not very efficient as it requires a lot of matches to make meaningful comparisons. With 4 years of data our calculation are still quite imprecise at the referee level due to the small number of matches for each referees. While the differences to the control group are quite large for most referees only one is statistically significant thanks to a number of matches large enough.

Another approach is to use the prediction of our model to identify matches and referees where the pattern of decisions shows more favouritism. Let N_A be the number of decisions taken for team A and N_B be the number of decisions taken for team B . Following

| Referee | Difference | t-stat | Nb match | N fixtures |
|---------|------------|--------|----------|------------|
| Ref 1 | -34.66 | -2.59 | 6 | 3 |
| Ref 2 | -33.22** | -4.72 | 12 | 4 |
| Ref 3 | -12.94 | -0.97 | 12 | 5 |
| Ref 4 | -5.52 | -0.79 | 18 | 7 |
| Ref 5 | -1.67 | -0.24 | 22 | 8 |
| Ref 6 | 3.88 | 0.74 | 12 | 5 |

Significant at ** 5%.

Table 4: Comparing referees individually to a control group (English SL referees vs Australian referees on similar fixtures)

Proposition 8, we have:

$$\mathbb{E} \left[\sum_{k=1}^{N_A} \frac{1}{N_A} \frac{d\Delta_{\beta_d} P}{d\beta_d} \Big|_{d=1} \right] > \mathbb{E} \left[\sum_{k=1}^{N_B} \frac{1}{N_B} \frac{d\Delta_{\beta_d} P}{d\beta_d} \Big|_{d=-1} \right]$$

With the addition of Proposition 1, we have:

$$\mathbb{E} \left[\sum_{k=1}^{N_A} \frac{d\Delta_{\beta_d} P}{d\beta_d} \Big|_{d=1} \right] > \mathbb{E} \left[\sum_{k=1}^{N_B} \frac{d\Delta_{\beta_d} P}{d\beta_d} \Big|_{d=-1} \right]$$

A simple way to use these predictions is to measure either the difference in the average impact of decisions:

$$D_1 = \sum_{k=1}^{N_A} \frac{1}{N_A} \frac{d\Delta_{\beta_d} P}{d\beta_d} \Big|_{d=1} - \sum_{k=1}^{N_B} \frac{1}{N_B} \frac{d\Delta_{\beta_d} P}{d\beta_d} \Big|_{d=-1}$$

Or, taking into account the number of decisions taken, the overall effect on the match outcome of the decision process:

$$D_2 = \sum_{k=1}^{N_A} \frac{d\Delta_{\beta_d} P}{d\beta_d} \Big|_{d=1} - \sum_{k=1}^{N_B} \frac{d\Delta_{\beta_d} P}{d\beta_d} \Big|_{d=-1}$$

Both these measures are imperfect to measure favouritism but can be seen as a decision tool to help identify the situations where more screening and review of decisions are required. These index cannot be seen as perfect measures of the effect of favouritism on match as only part of the consequence of favouritism are measured with these index. Both index have by design some limitation: the D_1 index (average difference in the value of decision) is useful it does not take into account the number of decisions, so it may underestimate favouritism, the D_2 index takes into account the number of decisions but number of decision may well be due to the natural game itself and as a consequence it may overestimate or underestimate favouritism. Table 5 shows the results of such an estimations. The fairness index allows to identify three referees in our sample of 7 with a fairness index significantly different from neutrality. Interestingly this procedure requires much less matches but information on decision within matches. The test of significance of the index identifies situations where the timing of decisions has given an advantage to a team beyond what should be expected from pure randomness. It should not be seen as proving the favouritism of the concerned referees, but should be used as a way to screen referees' performance and matches where referees decision may have been more atypical.

| Referee | D1 | D2 | Nb of matches |
|---------|-------------------|---------------------|---------------|
| Ref 1 | 3.25*** | 107.20*** | 1 |
| Ref 2 | 2.56 | 71.79 | 1 |
| Ref 3 | 2.26 [†] | 133.32 [†] | 2 |
| Ref 4 | 1.52 [†] | 92.62 [†] | 2 |
| Ref 5 | 0.22 | 15.58 | 2 |
| Ref 6 | -0.15 | -5.59 | 1 |
| Ref 7 | -1.01 | -34.21 | 1 |

Significant at [†] 10%, *** 1%. Test by non parametric bootstrap.

Positive numbers suggest favouritism toward English teams.

Table 5: Fairness index - timings of decision and impact on match outcome

6 Conclusion

This study has given for the first evidence of referees' national favouritism in rugby does exist and that this issue should not be ignored by sporting body. While we identify referee favouritism in the clear situation of nationality differences, it raise clearly the question of favouritism when referees may come from the same region or city of a team in a national league. In addition, favouritism could more subtle with referees having a preference on the outcome of the game as in situations where a team A can benefit from team B beating team C. Such situations are common in round tournaments. In rugby, the Tri-nation and Six-nation tournaments could be concerned by such an issue even if there is never a non neutral refereeing.

We show in this paper that how the technique used to identify favouritism could be used to measure a fairness index for matches and referees. Such an index could be a tool for sporting federations to monitor more closely referees decisions when neutrality may be questionable. This tool would increase the cost of biased decisions by revealing atypical patterns in referees decisions. As such it would most likely reduce the degree of favouritism by its simple existence. Such a tools could also be used by federation to help identify match fixing situations as those which have appeared recently in football in Germany and Italy.

A Technical appendix

Proof of Proposition 3. The left expression in equation (1) reaches a maximum for $-\Delta_t - \theta\phi > 0$, that is when the scoreline is close or when team A is slightly behind. Let consider

the situation $\phi = 0$ where both teams have equal ability.

$$\delta(\Delta_t) = \Phi\left(\frac{-\Delta_t}{\theta\sigma}\right) - \Phi\left(\frac{-\Delta_t - \beta_{d_t}}{\theta\sigma}\right)$$

$\delta(\Delta_t)$ is maximum for

$$\delta'(\Delta_t) = \frac{1}{\theta\sigma} \left(-\varphi\left(\frac{-\Delta_t}{\theta\sigma}\right) + \varphi\left(\frac{-\Delta_t - \beta_{d_t}}{\theta\sigma}\right) \right) = 0$$

As $\varphi\left(\frac{-\Delta_t}{\theta\sigma}\right) = -\varphi\left(\frac{\Delta_t}{\theta\sigma}\right)$, $\delta(\Delta_t)$ is maximum for: $\Delta_t = -\Delta_t - \beta_{d_t}$. Which gives: $\Delta_t = -\frac{\beta_{d_t}}{2} < 0$. \square

Proof of Proposition 6. In the case of two teams with identical ability, $\phi = 0$. The marginal effect of a decision on the winning probability of team A is:

$$\delta(\Delta_t, \theta, \beta_d) = \Phi\left(\frac{-\Delta_t}{\theta\sigma}\right) - \Phi\left(\frac{-\Delta_t - \beta_{d_t}}{\theta\sigma}\right)$$

We have

$$\delta'_{\beta_d}(\Delta_t, \theta, \beta_d) = \frac{1}{\theta\sigma} \varphi\left(\frac{-\Delta_t - \beta_{d_t}}{\theta\sigma}\right)$$

$\delta''_{\beta_d, \theta}(\Delta_t, \theta, \beta_d)$ is of the sign of $(-\theta^2\sigma^2 + \Delta^2 + 2\Delta\beta + \beta^2)$. That is $\delta''_{\beta_d, \theta}(\Delta_t, \theta, \beta_d) > 0$ iff $\theta < \frac{(\Delta + \beta)}{\sigma}$. Therefore, for $\Delta > \theta\sigma - \beta$, $\delta''_{\beta_d, \theta}(\Delta_t, \theta, \beta_d) > 0$. That is, for Δ large enough, the marginal effect of the decision increases with the time left until the end of the game (eg. it decreases over time). Reversely, for $\Delta < \theta\sigma - \beta$, $\delta''_{\beta_d, \theta}(\Delta_t, \theta, \beta_d) < 0$. That is, for Δ small enough, the marginal effect of the decision decreases with the time left until the end of the game (eg. it increases over time). \square

Proof of Proposition 7. Let call $\Delta_{\beta_d}P = \Phi\left(\frac{-\Delta_t - \theta\phi}{\theta\sigma}\right) - \Phi\left(\frac{-\Delta_t - \theta\phi - \beta_{d_t}}{\theta\sigma}\right)$. When $\sum_t^{t-1} d_t < 0$, the agent makes a biased decision for team A if $\Delta_{\beta_d} > a_t - \psi$. When $\sum_t^{t-1} d_t > 0$, the agent makes a biased decision for team A if $\Delta_{\beta_d} > a_t + \psi$. A biased decision for team B is however optimal if: $\psi - \Delta_{\beta_d} > a_t$. \square

Proof of Proposition 8. Biased decisions for team A are taken when $\left|\frac{d\Delta_{\beta_d}P}{d\beta_d}\right|_{d=1} > a + \psi$ while biased decision for team B are taken when $\left|\frac{d\Delta_{\beta_d}P}{d\beta_d}\right|_{d=-1} < a - \psi$. Consequently,

$$\left|\frac{d\Delta_{\beta_d}P}{d\beta_d}\right|_{d=1} > \left|\frac{d\Delta_{\beta_d}P}{d\beta_d}\right|_{d=-1} \quad \square$$

References

- ANSORGE, C., AND J. SCHEER (1988): “International Bias Detected in Judging Gymnastic Competition at the 1984 Olympic Games,” *Research Quarterly for Exercise and Sport*, 59(2), 103–07.
- ARIELY, D. (2008): *Predictably irrational: The hidden forces that shape our decisions*. HarperCollins London.
- BOYKO, R., A. BOYKO, AND M. BOYKO (2007): “Referee bias contributes to home advantage in English Premiership football,” *Journal of Sports Sciences*, 25(11), 1185–1194.
- BURAIMO, B., D. FORREST, AND R. SIMMONS (2007): “The twelfth man? Refereeing bias in English and German soccer,” *Governance and Competition in Professional Sports Leagues*, p. 217.
- CAMPBELL, B., AND J. GALBRAITH (1996): “Nonparametric tests of the unbiasedness of Olympic figure-skating judgments,” *The Statistician*, 45(4), 521–526.
- CROWE, S., AND J. MIDDELDORP (1996): “A comparison of leg before wicket rates between Australians and their visiting teams for test cricket series played in Australia, 1977-94,” *Journal of the Royal Statistical Society. Series D (The Statistician)*, 45(2), 255–262.
- DAWSON, P., S. DOBSON, J. GODDARD, AND J. WILSON (2007): “Are football referees really biased and inconsistent?: evidence on the incidence of disciplinary sanction in the English Premier League,” *Journal of the Royal Statistical Society-Series A*, 170(1), 231–250.
- DISTASO, W., L. LEONIDA, D. PATTI, AND P. NAVARRA (2008): “Corruption and Referee Bias in Football: the Case of Calciopoli,” Working paper.
- DOHMEN, T. (2008): “The influence of social forces: evidence from the behavior of football referees,” *Economic Inquiry*, 46(3), 411–424.
- EMERSON, J., M. SELTZER, AND D. LIN (2009): “Assessing Judging Bias: An Example From the 2000 Olympic Games,” *The American Statistician*, 63(2), 124–131.
- GARICANO, L., I. PALACIOS-HUERTA, AND C. PRENDERGAST (2005): “Favoritism under social pressure,” *Review of Economics and Statistics*, 87(2), 208–216.
- GREENWALD, A., D. MCGHEE, J. SCHWARTZ, ET AL. (1998): “Measuring individual differences in implicit cognition: The implicit association test,” *Journal of personality and social psychology*, 74, 1464–1480.

-
- JENDRUSCH, G. (2002): "Probleme bei der Bewegungsbeobachtung und-beurteilung durch Kampf-, Schieds-und Linienrichter," *Psychologie und Sport*, 9(4), 133–144.
- JONES, P. (2002): "Erskine (2002) MV Jones, GC Paull and J. Erskine, The impact of a team's reputation on the decisions of association football referees," *Journal of Sports Sciences*, 20, 991–1000.
- LEHMAN, D., AND A. REIFMAN (1987): "Spectator influence on basketball officiating," *The Journal of social psychology*, 127(6), 673–675.
- MCCONNELL, A., AND J. LEIBOLD (2001): "Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes," *Journal of Experimental Social Psychology*, 37(5), 435–442.
- MOHR, P., AND K. LARSEN (1998): "Ingroup favoritism in umpiring decisions in Australian football," *Journal of Social Psychology*, 138, 495–504.
- MYERS, T., N. BALMER, A. NEVILL, AND Y. AL-NAKEEB (2006): "Evidence of nationalistic bias in MuayThai," *Journal of Sports Science and Medicine*.
- NEVILL, A., N. BALMER, AND A. WILLIAMS (2002): "The influence of crowd noise and experience upon refereeing decisions in football," *Psychology of Sport & Exercise*, 3(4), 261–272.
- PAGE, K., AND L. PAGE (2009): "Alone against the crowd: Individual differences in referees' ability to cope under pressure," *Journal of Economic Psychology*.
- PLESSNER, H., AND T. HAAR (2006): "Sports performance judgments from a social cognitive perspective," *Psychology of Sport & Exercise*, 7(6), 555–575.
- POPOVIĆ, R. (2000): "International bias detected in judging rhythmic gymnastics competition at Sydney-2000 Olympic Games," *Physical Education and Sport*, 1(7), 1–3.
- RAINEY, D., J. LARSEN, AND A. STEPHENSON (1989): "The effects of a pitcher's reputation on umpires' calls of balls and strikes," *Journal of Sport Behavior*, 12, 139–150.
- RICKMAN, N., AND R. WITT (2008): "Favoritism and Financial Incentives: A Natural Experiment," *Economica*, 75(298), 296–309.
- SALMELA, J. (1978): "Gymnastics judging: A complex information processing task, or (who's putting one over on who?) Part 1," *International Gymnast*, 20, 54–56.
- SCOPPA, V. (2008): "Are subjective evaluations biased by social factors or connections? An econometric analysis of soccer referee decisions," *Empirical Economics*, 35(1), 123–140.

-
- SELTZER, R., AND W. GLASS (1991): “International politics and judging in Olympic skating events: 1968–1988,” *Journal of Sport Behavior*, 14, 189–200.
- STE-MARIE, D. (1996): “International Bias in Gymnastic Judging: Conscious or Unconscious Influences?,” *Perceptual and motor skills*, 83(3), 963–975.
- SUTTER, M., AND M. KOCHER (2004): “Favoritism of agents—The case of referees’ home bias,” *Journal of Economic Psychology*, 25(4), 461–469.
- WHISSELL, R., S. LYONS, D. WILKINSON, AND C. WHISSELL (1993): “National bias in judgments of Olympic-level skating,” *Perceptual and motor skills*, 77(2), 355–358.
- ZITZEWITZ, E. (2006): “Nationalism in winter sports judging and its lessons for organizational decision making,” *Journal of Economics and Management Strategy*, 15(1), 67.

List of NCER Working Papers

No. 61 ([Download full text](#))

Nicholas King, P Dorian Owen and Rick Audas

Playoff Uncertainty, Match Uncertainty and Attendance at Australian National Rugby League Matches

No. 60 ([Download full text](#))

Ralf Becker, Adam Clements and Robert O'Neill

A Cholesky-MIDAS model for predicting stock portfolio volatility

No. 59 ([Download full text](#))

P Dorian Owen

Measuring Parity in Sports Leagues with Draws: Further Comments

No. 58 ([Download full text](#))

Don Harding

Applying shape and phase restrictions in generalized dynamic categorical models of the business cycle

No. 57 ([Download full text](#))

Renee Fry and Adrian Pagan

Sign Restrictions in Structural Vector Autoregressions: A Critical Review

No. 56 ([Download full text](#))

Mardi Dungey and Lyudmyla Hvozdyk

Cojumping: Evidence from the US Treasury Bond and Futures Markets

No. 55 ([Download full text](#))

Martin G. Kocher, Marc V. Lenz and Matthias Sutter

Psychological pressure in competitive environments: Evidence from a randomized natural experiment: Comment

No. 54 ([Download full text](#))

Adam Clements and Annastiina Silvennoinen

Estimation of a volatility model and portfolio allocation

No. 53 ([Download full text](#))

Luis Catão and Adrian Pagan

The Credit Channel and Monetary Transmission in Brazil and Chile: A Structured VAR Approach

No. 52 ([Download full text](#))

Vlad Pavlov and Stan Hurn

Testing the Profitability of Technical Analysis as a Portfolio Selection Strategy

No. 51 ([Download full text](#))

Sue Bridgewater, Lawrence M. Kahn and Amanda H. Goodall

Substitution Between Managers and Subordinates: Evidence from British Football

No. 50 ([Download full text](#))

Martin Fukac and Adrian Pagan

Structural Macro-Econometric Modelling in a Policy Environment

No. 49 ([Download full text](#))

Tim M Christensen, Stan Hurn and Adrian Pagan

Detecting Common Dynamics in Transitory Components

No. 48 ([Download full text](#))

Egon Franck, Erwin Verbeek and Stephan Nüesch

Inter-market Arbitrage in Sports Betting

No. 47 ([Download full text](#))

Raul Caruso

Relational Good at Work! Crime and Sport Participation in Italy. Evidence from Panel Data Regional Analysis over the Period 1997-2003.

No. 46 ([Download full text](#)) (Accepted)

Peter Dawson and Stephen Dobson

The Influence of Social Pressure and Nationality on Individual Decisions: Evidence from the Behaviour of Referees

No. 45 ([Download full text](#))

Ralf Becker, Adam Clements and Christopher Coleman-Fenn

Forecast performance of implied volatility and the impact of the volatility risk premium

No. 44 ([Download full text](#))

Adam Clements and Annastiina Silvennoinen

On the economic benefit of utility based estimation of a volatility model

No. 43 ([Download full text](#))

Adam Clements and Ralf Becker

A nonparametric approach to forecasting realized volatility

No. 42 ([Download full text](#))

Uwe Dulleck, Rudolf Kerschbamer and Matthias Sutter

The Economics of Credence Goods: On the Role of Liability, Verifiability, Reputation and Competition

No. 41 ([Download full text](#))

Adam Clements, Mark Doolan, Stan Hurn and Ralf Becker

On the efficacy of techniques for evaluating multivariate volatility forecasts

No. 40 ([Download full text](#))

Lawrence M. Kahn

The Economics of Discrimination: Evidence from Basketball

No. 39 ([Download full text](#))

Don Harding and Adrian Pagan

An Econometric Analysis of Some Models for Constructed Binary Time Series

No. 38 ([Download full text](#))

Richard Dennis

Timeless Perspective Policymaking: When is Discretion Superior?

No. 37 ([Download full text](#))

Paul Frijters, Amy Y.C. Liu and Xin Meng

Are optimistic expectations keeping the Chinese happy?

No. 36 ([Download full text](#))

Benno Torgler, Markus Schaffner, Bruno S. Frey, Sascha L. Schmidt and Uwe Dulleck

Inequality Aversion and Performance in and on the Field

No. 35 ([Download full text](#))

T M Christensen, A. S. Hurn and K A Lindsay

Discrete time-series models when counts are unobservable

No. 34 ([Download full text](#))

Adam Clements, A S Hurn and K A Lindsay

Developing analytical distributions for temperature indices for the purposes of pricing temperature-based weather derivatives

No. 33 ([Download full text](#))

Adam Clements, A S Hurn and K A Lindsay

Estimating the Payoffs of Temperature-based Weather Derivatives

No. 32 ([Download full text](#))

T M Christensen, A S Hurn and K A Lindsay

The Devil is in the Detail: Hints for Practical Optimisation

No. 31 ([Download full text](#))

Uwe Dulleck, Franz Hackl, Bernhard Weiss and Rudolf Winter-Ebmer

Buying Online: Sequential Decision Making by Shopbot Visitors

No. 30 ([Download full text](#))

Richard Dennis

Model Uncertainty and Monetary Policy

No. 29 ([Download full text](#))

Richard Dennis

The Frequency of Price Adjustment and New Keynesian Business Cycle Dynamics

No. 28 ([Download full text](#))

Paul Frijters and Aydogan Ulker

Robustness in Health Research: Do differences in health measures, techniques, and time frame matter?

No. 27 ([Download full text](#))

Paul Frijters, David W. Johnston, Manisha Shah and Michael A. Shields

Early Child Development and Maternal Labor Force Participation: Using Handedness as an Instrument

No. 26 ([Download full text](#))

Paul Frijters and Tony Beaton

The mystery of the U-shaped relationship between happiness and age.

No. 25 ([Download full text](#))

T M Christensen, A S Hurn and K A Lindsay

It never rains but it pours: Modelling the persistence of spikes in electricity prices

No. 24 ([Download full text](#))

Ralf Becker, Adam Clements and Andrew McClelland

The Jump component of S&P 500 volatility and the VIX index

No. 23 ([Download full text](#))

A. S. Hurn and V. Pavlov

Momentum in Australian Stock Returns: An Update

No. 22 ([Download full text](#))

Mardi Dungey, George Milunovich and Susan Thorp

Unobservable Shocks as Carriers of Contagion: A Dynamic Analysis Using Identified Structural GARCH

No. 21 ([Download full text](#)) (forthcoming)

Mardi Dungey and Adrian Pagan

Extending an SVAR Model of the Australian Economy

No. 20 ([Download full text](#))

Benno Torgler, Nemanja Antic and Uwe Dulleck

Mirror, Mirror on the Wall, who is the Happiest of Them All?

No. 19 ([Download full text](#))

Justina AV Fischer and Benno Torgler

Social Capital And Relative Income Concerns: Evidence From 26 Countries

No. 18 ([Download full text](#))

Ralf Becker and Adam Clements

Forecasting stock market volatility conditional on macroeconomic conditions.

No. 17 ([Download full text](#))

Ralf Becker and Adam Clements

Are combination forecasts of S&P 500 volatility statistically superior?

No. 16 ([Download full text](#))

Uwe Dulleck and Neil Foster

Imported Equipment, Human Capital and Economic Growth in Developing Countries

No. 15 ([Download full text](#))

Ralf Becker, Adam Clements and James Curchin

Does implied volatility reflect a wider information set than econometric forecasts?

No. 14 ([Download full text](#))

Renee Fry and Adrian Pagan

Some Issues in Using Sign Restrictions for Identifying Structural VARs

No. 13 ([Download full text](#))

Adrian Pagan

Weak Instruments: A Guide to the Literature

No. 12 ([Download full text](#))

Ronald G. Cummings, Jorge Martinez-Vazquez, Michael McKee and Benno Torgler

Effects of Tax Morale on Tax Compliance: Experimental and Survey Evidence

No. 11 ([Download full text](#))

Benno Torgler, Sascha L. Schmidt and Bruno S. Frey

The Power of Positional Concerns: A Panel Analysis

No. 10 ([Download full text](#))

Ralf Becker, Stan Hurn and Vlad Pavlov

Modelling Spikes in Electricity Prices

No. 9 ([Download full text](#))

A. Hurn, J. Jeisman and K. Lindsay

Teaching an Old Dog New Tricks: Improved Estimation of the Parameters of Stochastic Differential Equations by Numerical Solution of the Fokker-Planck Equation

No. 8 ([Download full text](#))

Stan Hurn and Ralf Becker

Testing for nonlinearity in mean in the presence of heteroskedasticity.

No. 7 ([Download full text](#)) (published)

Adrian Pagan and Hashem Pesaran

On Econometric Analysis of Structural Systems with Permanent and Transitory Shocks and Exogenous Variables.

No. 6 ([Download full text](#)) (published)

Martin Fukac and Adrian Pagan

Limited Information Estimation and Evaluation of DSGE Models.

No. 5 ([Download full text](#))

Andrew E. Clark, Paul Frijters and Michael A. Shields

Income and Happiness: Evidence, Explanations and Economic Implications.

No. 4 ([Download full text](#))

Louis J. Maccini and Adrian Pagan

Inventories, Fluctuations and Business Cycles.

No. 3 ([Download full text](#))

Adam Clements, Stan Hurn and Scott White

Estimating Stochastic Volatility Models Using a Discrete Non-linear Filter.

No. 2 ([Download full text](#))

Stan Hurn, J.Jeisman and K.A. Lindsay

Seeing the Wood for the Trees: A Critical Evaluation of Methods to Estimate the Parameters of Stochastic Differential Equations.

No. 1 ([Download full text](#))

Adrian Pagan and Don Harding

The Econometric Analysis of Constructed Binary Time Series.